

Estimating the Partisan Bias of Survey Questions

Aaron R. Kaufman Harvard University <http://www.scholar.harvard.edu/aaronkaufman/>

Abstract I introduce an efficient automated method for estimating the partisan bias in survey questions which arises from question wording, then validate the model experimentally. I also introduce an application in R Shiny to estimate the bias in user-inputted questions.

Method

Training Set

Congressional Record 1994-2014
8 million docs
+
Ideal points of every member of Congress

Classifier Ensemble →

Model

Word	Bias	Word	Bias
Moderate	-0.235	Quick	+0.522
Barrier	-0.217	Pressure	+0.466
Illegal Immigrant	-0.214	Earmark	+0.428
Undocumented	-0.210	Angry	+0.359
Trust	-0.160	Sport	+0.343
Discrimination	-0.155	Energy Technology	+0.304

Test Set

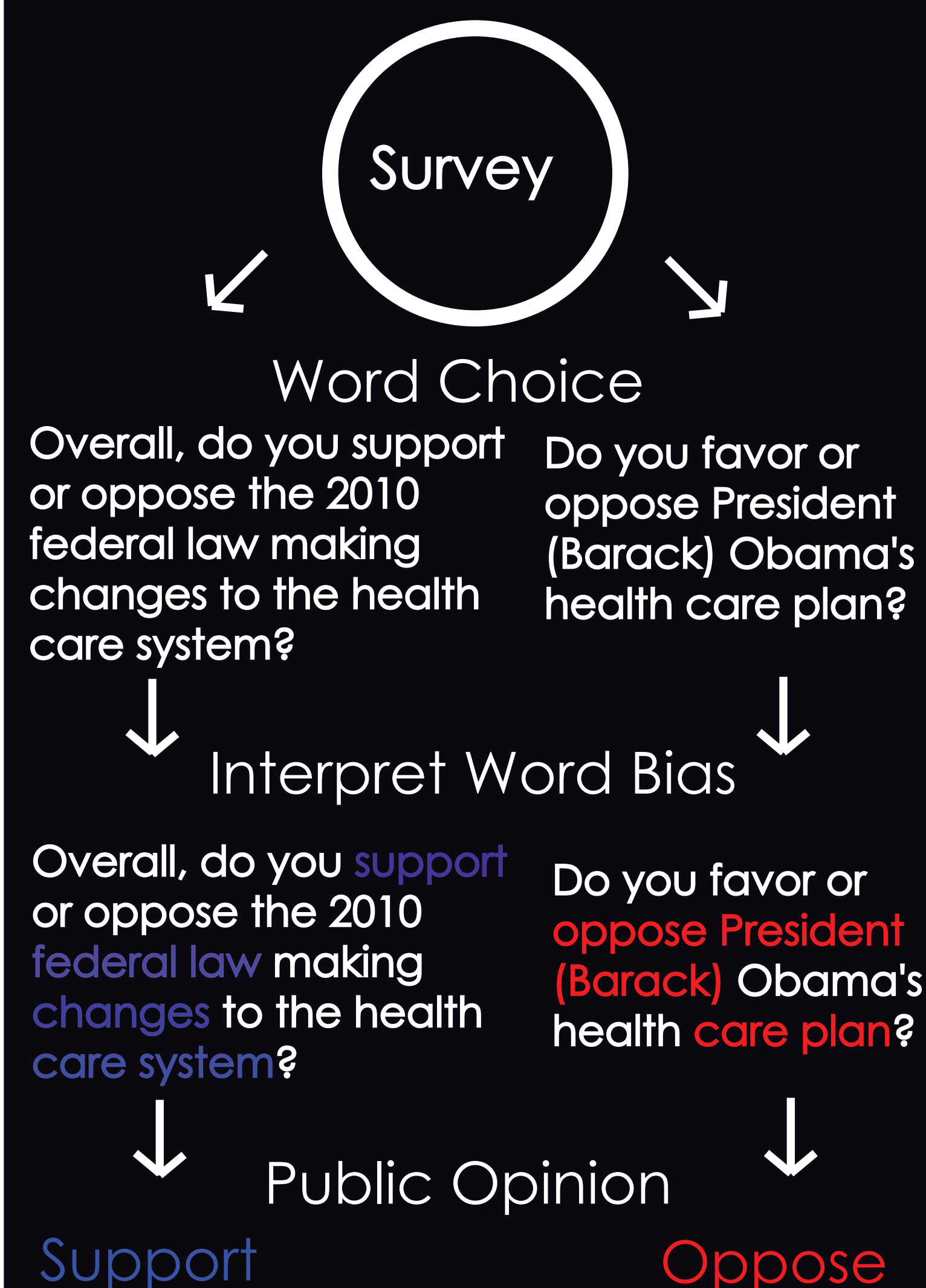
Roper Center's iPoll Database 2009-2014
800,000 survey questions

→ Trained Model →

Question Bias

Which should be the higher priority now—securing the nation's border, or addressing the status of illegal immigrants currently in the US?	-1.150
How much discrimination against blacks do you feel there is in the United States today, limiting their chances to get ahead?	-0.028
Are you sometimes afraid that the police will stop and arrest you when you are completely innocent, or not?	+0.410
Are we spending too much or too little on space exploration programs?	+1.325

Outline



Measurement

"Are we spending too much or too little on space exploration programs?" Survey Question

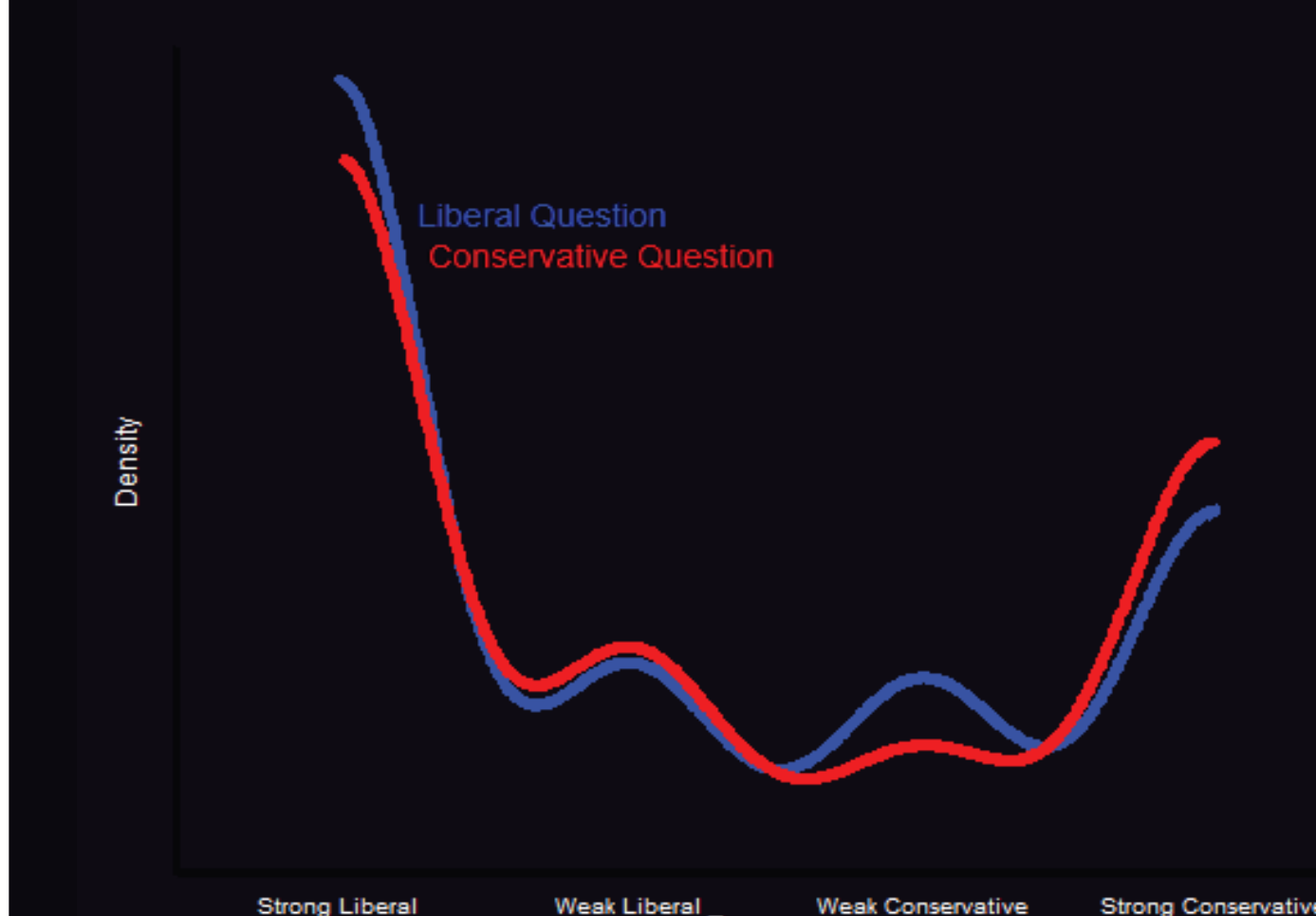
Spend spendmuch muchlittl littl littl- space space spaceexplor explor explorprogram program

Clean, stem
Featurize into
unigrams and
bigrams

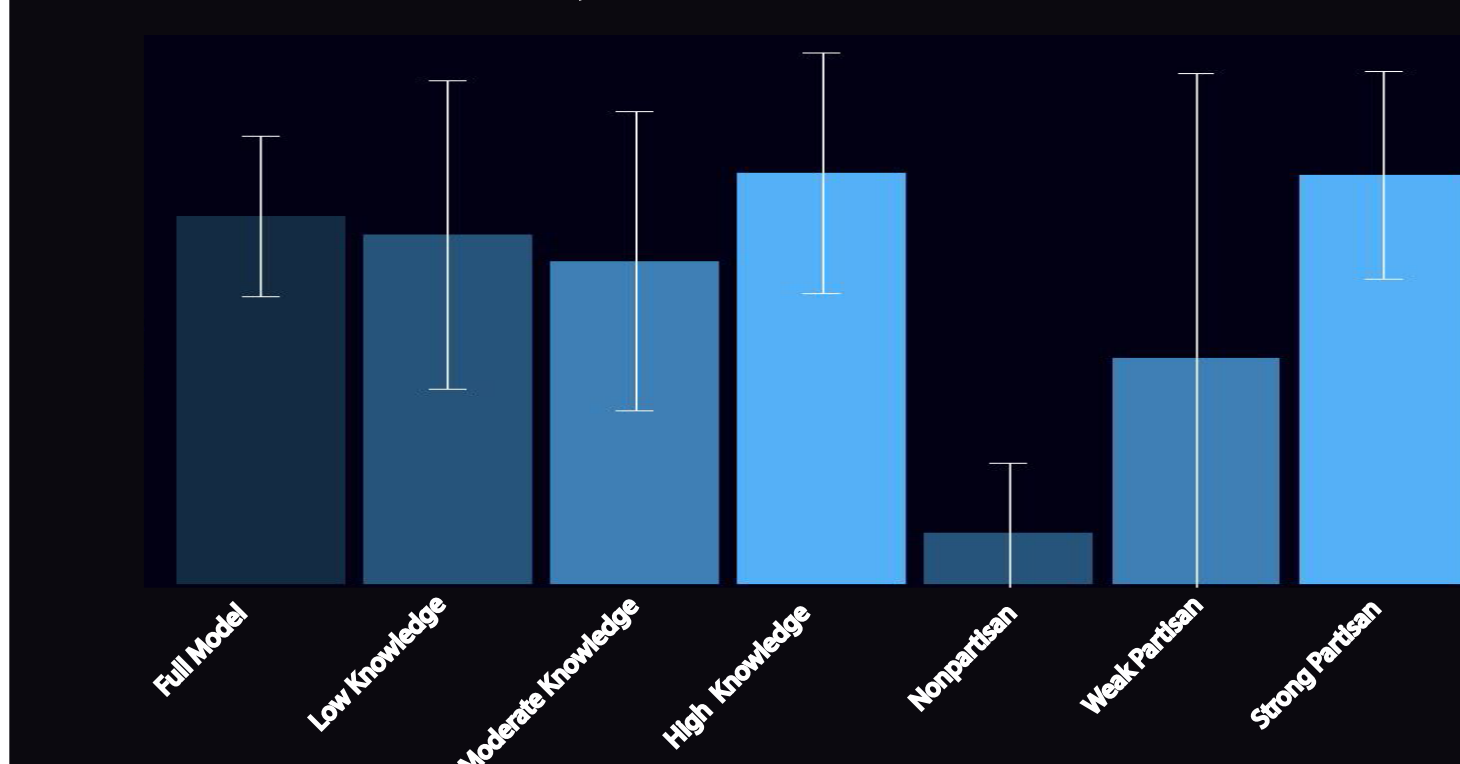


Validation

Liberal prompts produce liberal responses



Partisans and sophisticates respond more strongly to lexical cues



bit.ly/1mcGVHi

This app takes any unstructured text and returns the partisan bias, keyword contributions, and the nearest senator by ideal point

Survey Question Bias Estimator: Start by inputting some text in the box on the left!

Input your survey question below:

Do you approve or strongly approve of this application?

Select a training corpus below:

Congressional Record 2009-2014

Keywords: Bias, Keywords, Nearest Senator, How to Cite

Your question's estimated bias is:

(1) 0.4670888

And the 95% confidence interval around that estimate is:

(1) 0.0344878 0.0402834

This bias score ranges from -2, extremely liberal, to +2, extremely conservative.

Survey Question Bias Estimator: Start by inputting some text in the box on the left!

